African Invertebrates 65(2): 223–246 (2024)
DOI: 10.3897/AfrInvertebr.65.138881

Research Article

Completing the web: identifying sampling bias and knowledge gaps within South African spider surveys (Arachnida, Araneae)

Aileen C. van der Mescht[1], Charles R. Haddad[1], Stefan H. Foord[2†], Ansie S. Dippenaar-Schoeman[2]

1 *University of the Free State, Bloemfontein, South Africa*
2 *University of Venda, Thohoyandou, South Africa*
Corresponding author: Aileen C. van der Mescht (aileenvandermescht@gmail.com)

Abstract

Species distribution datasets are fundamental for macroecological studies, although there is an overarching need to ensure that these datasets are representative of the entire community. Shortfalls, or knowledge gaps, within biodiversity datasets originate for a range of reasons, and can lead to incorrect conclusions or recommendations being drawn. Spatial scale influences the interpretations of diversity patterns and thus is an important aspect to consider. South Africa has a rich history of spider sampling and as such, it is possible to investigate the influence that scale, both spatial and taxonomic, has on the overall interpretations of how complete the spider knowledge base is in the country. To do this, we draw on curated natural history spider collections and determine how complete the spider assemblages are across twelve unique combinations of taxonomic and spatial scales. Overall, we received 121 605 usable records from seven collections, with spider records and diversity, being concentrated along the eastern and coastal regions of South Africa. We show that assemblage completeness increases with both increasing taxonomic and spatial scales, and as such, knowledge of the distribution of spider families at the biome level is largely complete. Moreover, we show that our fine-scale knowledge of spider assemblages in South Africa is relatively poor, yet we do identify, even at fine scales, assemblages in South Africa that can be considered complete. We identify under-sampled regions of the country, which in turn are congruent with the distribution of under-sampled regions found in other South African invertebrate groups. We show that the scaling of completeness can only be interpreted in one direction: as scale increases so does completeness. These findings will have important implications for spider research and conservation in South Africa, given that regions where completeness is highest correspond strongly to areas in South Africa with the highest threats to biodiversity.

Key words: Museum records, spatial scale, species accumulation, taxonomic scale

OPEN ACCESS

This article is part of:
Gedenkschrift for Prof. Stefan H. Foord
Edited by Galina Azarkina, Ansie Dippenaar-Schoeman, Charles Haddad, Robin Lyle, John MIdgley, Caswell Munyai

Academic editor: John Midgley
Received: 9 October 2024
Accepted: 19 November 2024
Published: 18 December 2024

ZooBank: https://zoobank.org/9EB46EAD-718C-493C-A4D9-8FB713112730

Citation: van der Mescht AC, Haddad CR, Foord SH, Dippenaar-Schoeman AS (2024) Completing the web: identifying sampling bias and knowledge gaps within South African spider surveys (Arachnida, Araneae). African Invertebrates 65(2): 223–246. https://doi.org/10.3897/AfrInvertebr.65.138881

Copyright: © Aileen C. van der Mescht et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

Species distribution datasets form the fundamental building blocks for many macroecological studies (Wüest et al. 2020; Cornford et al. 2021). Primary data sourced from natural history collections provide extremely valuable biodiversity information (Robertson et al. 2010; Scoble 2010), with much of the information

† Deceased.

regarding South Africa's biodiversity being provided by the country's natural history collections (Drinkrow et al. 1994; Hamer 2012). Beyond natural history collections, data can be sourced from published taxonomic descriptions and checklists, to citizen science databases such as iNaturalist (Dippenaar-Schoeman et al. 2012; Callaghan et al. 2020; Wolf et al. 2022; Garretson et al. 2023). Although the quality and reliability of curated versus untrained citizen science data can be debated (Aceves-Bueno et al. 2017; Jacobs and Zipf 2017; Fraisl et al. 2022), there is an overarching need to ensure that all datasets used in macroecological studies are a representative whole of the community (Qian 2020; Kusumoto et al. 2023; Alves-Martins et al. 2024). Sampling bias impacting species to geographic regions of interest will result in inherent biases within databases (Dippenaar-Schoeman et al. 2012), and thus unbiased interpretations based on these data are impossible (Yang et al. 2013).

Within large biodiversity datasets, various shortfalls regarding the completeness of these databases exist (Hortal et al. 2015). Inevitably named after prominent taxonomists or ecologists, the main shortfalls encountered in many datasets impact a range of biodiversity aspects. Arguably, the two most apparent shortfalls are firstly the Linnean, where most of the species on earth are neither described nor catalogued, and secondly, the Wallacean shortfall, which refers to the fact that the geographic and temporal distribution of many species is incomplete (Hortal et al. 2015). Wallacean shortfalls have been identified in many datasets (Mora et al. 2008; Yang et al. 2013; Troia and McManamay 2016). In many cases, though, incomplete datasets for a region are because of a combination of both Wallacean and Linnean shortfalls, with knowledge gaps representing both the lack of distributional knowledge of a species, as well as the presence of undescribed species at the site (Oliveira et al. 2016).

These knowledge gaps and biodiversity shortfalls can result from a plethora of reasons (Whittaker et al. 2005; Foord et al. 2011a, 2011b; Hortal et al. 2015; Oliveira et al. 2016; Ramírez et al. 2022; Vergara-Asenjo et al. 2023). Beyond the obvious constraints regarding the time taken to collect, store, sort and identify samples (Cardoso et al. 2011; Foord et al. 2013; Janion-Scheepers et al. 2016; Wilkinson et al. 2021), the idiosyncrasies and personal preferences of "unbiased" collectors can result in geographical bias within databases. This occurs either when research or collections are undertaken at preferential sites, such as nature reserves and scenic areas (Sánchez-Fernández et al. 2022), or close to access routes, or when higher rates of sampling occur in regions expected to be more diverse (Oliveira et al. 2016), such as the global biodiversity hotspots (Myers et al. 2000). Furthermore, in many cases sampling locations are highly correlated to the locations of research institutes and universities (Oliveira et al. 2016; Sánchez-Fernández et al. 2022). As such, species richness estimators and species accumulation curves are traditionally used to identify regions where sampling is complete (Chao and Jost 2012; Chao et al. 2020).

Diversity patterns, and the interpretations thereof, are highly influenced by scale (Whittaker et al. 2005; Foord et al. 2008), with challenges arising when attempting to extrapolate from one scale to another (Teng et al. 2020). Alpha diversity (α) describes diversity at a local scale, whereas beta (ß) and gamma (γ) diversity describe the turnover of diversity between sites, and thus both describe diversity across larger scales (Burley et al. 2016; Foord and Dippenaar-Schoeman

2016). At a local scale, species responses to the landscape can be modulated either by local factors, such as the presence of suitable habitats leading to habitat filtration of species (Pärtel et al. 2016), or by biogeographical factors that occur across larger spatial scales (Banks-Leite et al. 2022). Thus, the interpretation of local and regional patterns of biodiversity may vary considerably depending on the scale at which patterns are observed (Pärtel et al. 2016; Banks-Leite et al. 2022), with variable functional relationships existing between ecological variables at different scales (Teng et al. 2020).

Variation in environmental drivers at local sites (Pärtel et al. 2016) and differing responses of species across different regions and biomes (Foord et al. 2011b; Haddad et al. 2013; Banks-Leite et al. 2022) leads to complex and varied interactions between factors at local and regional scales. For example, local α-diversity can be the result of local responses to habitat availability and filtration (Pärtel et al. 2016), yet this α-diversity is also dependent of the regional pool of species (ß- and γ-diversity), yet the degree of specialisation and diversity (or lack thereof) of species at a local site can also be as a result of interactions of various spatiotemporal factors, species dispersal abilities or even geographical barriers (Burley et al. 2016).

Scale is essential to consider when investigating the distribution of diversity, be it local or regional, as differing mechanisms may emerge as drivers of species distributions (Gómez-Rodríguez and Baselga 2018; Martín-Devasa et al. 2024). Considering spiders in particular, local richness and composition are positively driven by local ecological factors such as habitat heterogeneity (Clough et al. 2005; Jiménez-Valverde and Lobo 2007; De Mas et al. 2009; Haddad et al. 2019), site context and local landscape configuration (Clough et al. 2005) and vegetation complexity (Clough et al. 2005). Furthermore, habitat (between plant types) and microhabitat (within the same plant type) have been shown to variably impact the colonisation and specialisation (phylogenetic variation) of the associated spider communities. When considering the habitat level, spider size and shape are filtered, whereas spider evolutionary adaptations as well as size and shape are selected on at the microhabitat level (Gonçalves-Souza et al. 2014; Wilson et al. 2023).

When model performance is considered, explained variation, as well as the environmental variables identified, will vary across different local scales. Thus, a larger scale model may suggest an ecological variable of importance, that when applied to local conservation or management protocols may in fact be less appropriate and less effective at managing and protecting spider communities. For example, at a local scale spanning 230 × 230 m (the model with the highest explained variance) and all other smaller scales considered up to this point, the rock terrain, percentage sclerophyllous vegetation and the standard deviation of NDVI best predicted spider species richness (De Mas et al. 2009). Yet, at larger scales, the variation explained by the models decreased, as well as the number of significant explanatory variables, until only percentage sclerophyllous vegetation explained half the variance of the best model (De Mas et al. 2009). Furthermore, species-specific responses vary across different local scales (Schmidt et al. 2008).

The extrapolation of the local scale models attempting to predict the distribution of spider diversity to larger scales fail largely due to the lack of structural environmental variable(s) for the area (Jiménez-Valverde and

Lobo 2007; De Mas et al. 2009; Joseph et al. 2018). At a continental scale, spider assemblages across Europe are shaped by the dispersal limitation of individual species. Yet, when southern and northern Europe are considered as individual units (sub-continental scale), dispersal limitation is the principle shaping force of spider assemblages in southern Europe, while no causal force can be isolated for northern European spider assemblages (Martín-Devasa et al. 2024).

The history of spider sampling in South Africa spans over 300 years, with the first two spiders to be described in South Africa in the 1700's. From then on, the number of described species rapidly increased up until the early 1900's, with peak rates of descriptions up until the 1920's. This was followed by a relative slowing in the rate at which species were described until 1997 (Dippenaar-Schoeman et al. 2023). In 1997, the South African National Survey of Arachnida (SANSA) was initiated to determine the extent of the South African Arachnida biodiversity, as well as identify gaps in the geographic distribution of species (Dippenaar-Schoeman et al. 2015). Since then, the rate of species descriptions has increased remarkably (Dippenaar-Schoeman et al. 2023). The First Atlas of South African Spiders was published in 2010 (Dippenaar-Schoeman et al. 2010), which provided maps for 2010 species of spiders from 71 families. This geographical data was essential in preparing the first Red List of the South African spiders (Foord et al. 2020), as well as a National Spider Checklist of 2265 spider species (Dippenaar-Schoeman et al. 2023), an increase of 255 species from 2010. The checklist breaks down the distribution of species richness and number of records per province in South Africa (Dippenaar-Schoeman et al. 2023), with ascensions functioning as a proxy for sampling effort. Given that this latest checklist makes a start at describing knowledge gaps, and sampling bias within the spider distributions in South Africa, quantification of these gaps through empirical means is the logical next step.

Thus, this study aims to quantify sample completeness within the South African spider assemblages. To achieve this, our first objective is to compile as comprehensive a database as possible using curated data from various sources. Secondly, we determine how completeness of the spider database varies in accordance to both spatial and taxonomic scales. We quantify completeness by comparing estimated species richness to observed species richness across the scales, accounting for the number of specimen records. We vary spatial scale by using both arbitrarily defined geographical boundaries (quarter and degree grid cells), as well as ecological boundaries (bioregion and biome) to do this. Taxonomic scale is varied across species, genus, and family level. We hypothesise that at both finer spatial and taxonomic levels, completeness within the database will be relatively low given how diverse spiders are known to be. However, as the scales become coarser, we hypothesise that the completeness will increase as sampling effort per taxonomic/geographic region increases, effectively reducing the accumulation of diversity as the sample size increases. Beyond identifying areas to target for further sampling, this study forms the basis upon which further macroecological spider studies in South Africa can be built by identifying bias and possible shortfalls within the already existing data, thus ensuring that these shortfalls can be addressed and bias reduced in any other study going forward.

Methods

Data collection

The majority of spider specimen records for South Africa (n = 73 649) used for this study were sourced from the National Collection of Arachnida (NCA) at the Agricultural Research Council in Pretoria. We did not include any records from partial enclave of eSwatini. We received spider specimen accession records from local and international collections, namely the Albany Museum, Grahamstown (n = 1 77); National Museum, Bloemfontein (n = 16 061); KwaZulu-Natal Museum, Pietermaritzburg (n = 10 517); Iziko South African Museum, Cape Town (n = 9 763); Ditsong National Museum of Natural History, Pretoria (n = 5 837) and Royal Museum of Central Africa, Tervuren, Belgium (n = 4 001). Only specimen records where the spiders were identified to either species, genus or family level were retained. Subspecies identifications were absorbed into the species level identification.

To ensure that species names across all data sources were accurate and valid, and to prevent duplication where individual samples may have been identified with old nomenclature, all species names were validated against the Spider Checklist of South Africa (Dippenaar-Schoeman et al. 2023). Where species names did not match with the checklist, CH validated each name, correcting for nomenclatural changes as well as spelling mistakes using the World Spider Catalog (2024).

A fundamental requirement of the data is that each spider record needed to be georeferenced with complete coordinates. Thus, the number of records received across all data sources was always greater than the number of records that we retained. All samples were plotted in QGIS (QGIS Development Team 2020), and sites where the coordinates did not match the provided location information were identified. Where possible, these coordinates were corrected, either by swapping the × and y coordinate or where location information was specific enough, coordinates for the location were used. Otherwise, incorrectly georeferenced sample records were excluded from the final database. The final spider sample database was saved as a shapefile. It must be noted that older specimens and records often have rudimentary descriptions attached or lack detailed morphological descriptions or collecting data (e.g. described from South Africa, Transvaal or Cape), so could not be included. Furthermore, some species remain known from the original descriptions only and have never been resampled and definitively identified, and as their type material resides in international collections not included in this study, such species were also omitted.

The SANSA database used in the production of the national checklist include all the NCA records as well as all published species records from local and international collections from the taxonomic literature. Considering the issues discussed above, the restricted number of collection databases used in our study, and the exclusion of published records not in the seven institutional databases were used, our final database contains approximately 180 fewer species than were recorded from South Africa by Dippenaar-Schoeman et al. (2023).

Mapping

South Africa was divided at four different spatial scales ranging from fine to coarse, namely 1) quarter degree and 2) degree grid cells, and then at the 3) bioregion and 4) biome levels. QGIS was used to create the two grid cell levels,

South Africa was divided first by a grid measuring 0.25° × 0.25° (approximately 24 km longitudinally × 27 km latitudinally), thus dividing the country into 1995 quarter degree grid cells (QDGC), using the WGS84 projection for South Africa. Each cell was named according to convention. This procedure was repeated a second time, where grids measuring 1° × 1° (approximately 97 km longitudinally × 110 km latitudinally) divided the country into 151 degree grid cells (DGC). Here, each cell was named according to the latitude and longitude comprising the intersection of the top left corner of each cell.

Bioregion and biome levels were extracted from the vegetation map of South Africa (South African National Biodiversity Institute 2006), with bioregion being a finer scale than biome (Rutherford et al. 2005). The spider sample shapefile was then overlaid on the two grid cell layers, as well as the vegetation map in *R* (R Core Team 2020). To determine in which QDGC, DGC, bioregion and biome individual spider samples occurred, the *st_join* function in the *sf* package (Pebesma 2018) was used. All further analyses were conducted on the resultant spider database containing both the taxonomic information relating to each spider sample, as well as spatial information relating its location.

Statistical analysis

Sample completeness was determined at three taxonomic levels, namely species-, genus- and family-level identifications. Four spatial scales were used to delimit spider communities, two arbitrary (QDGC and DGC grids) and two ecological (bioregion and biome). Thus, there were 12 unique combinations between taxonomic and spatial levels across which sample completeness was determined. As such, the spider database was first split by spatial level, and input matrices were then generated where individual spider samples were rows and either species, genus or family names were columns, depending on the spatial-taxonomic combination being assessed.

Biodiversity sample completeness of a site or region is often estimated using a rarefied species accumulation curve and determining whether the curve reaches an asymptote (Chao et al. 2014). The issue here is that in many cases an eyeball measure of completeness is then employed to state that the asymptote has been reached or that the sample completeness is approaching the asymptote without directly determining the asymptotic value, or how far the curve is away from the asymptote. With an asymptotic value, which functions as an estimate of the expected species richness of a sample (S_{est}), as well as observed species richness value (S_{obs}), it is possible to determine the percentage completeness of a sample such that:

$$Percentage\ \ completeness = \frac{S_{obs}}{S_{est}} \times 100$$

To calculate the percentage completeness, we used the function *iNEXT* in the package *iNEXT* (Chao et al. 2014; Hsiech et al. 2020) to calculate the asymptotic values for each spatial-taxonomic scale combination. Incidence-based frequencies of each species at a spatial scale were used. We then calculated the percentage completeness for each unit based on these first-order species richness estimators. In addition to percentage completeness, we then calculated the number of accession records per spatial-taxonomic combination. As the accuracy of

extrapolation methods on small sample sizes is often inaccurate and can result in a high occurrence of false positives, the number of accession records allowed us to filter these false positives out. In addition, we also calculated the completeness of the species, genus and family records of South Africa as a whole.

We defined four threshold levels of total sampling completeness based on cut-off values for the percentage completeness and number of accession records per unit. Under-sampled regions had a percentage completeness < 0.6 and < 10 accession records. Moderately well-sampled regions had a percentage ≥ 0.6 but ≤ 0.75 and ≥ 10 accession records. Relatively well-sampled regions had a percentage completeness of ≥ 0.76 but ≤ 0.9 and ≥ 25 accession records. Finally, well-sampled regions had a percentage completeness of ≥ 0.91 and ≥ 50 accession records. Accession record cut-off values were based on Troia and McManamay (2016) who used these values as cutoffs for a range of groups such as amphibians, plants, insect and fishes. The percentage completeness cut-off values were based on Sánchez-Fernández et al. (2022). Yang et al. (2013) calculated the slope of the accumulation curve based on the last 10% of the data to determine how complete a sample is, with Troia and McManamay (2016) doing the same and then adding a third threshold class based on slope values, with the assumption that a slope closer to 0 is almost at its asymptote. However, as this method relies on fitting a linear function to a non-linear curve – a somewhat mathematically dubious approach – we chose to use percentage completeness and the number of accession records only.

To quantify changes in sample completeness across increasing taxonomic levels (species < genus < family), we used an ordinal regression approach with the calculated sample completeness (described above) as the response variables in the models. We modelled each spatial scale separately, as the spatial boundaries of the QDGCs, DGCs, bioregions and biomes do not correspond. We used cumulative link mixed effect models (CLMMs), where we ranked sample completeness as under-sampled < moderately well-sampled < relatively well-sampled < well-sampled. Taxonomic level and grid cell identity were used as the categorical fixed effect term and random effect of the model, respectively, as there were three repeated measures of sample completeness for each modelled spatial scale. For the QDGC and DGC models, two iterations were run, one excluding and one including all grid cells that contained no data as the lowest level of completeness (no data < under-sampled < moderately well-sampled < relatively well-sampled < well-sampled). This was done to ensure that interpretations were not biased by empty grid cells. Thus, in total, six individual models were run. All models were fitted with Laplace approximations using the *clmm* function in the *ordinal* (Christensen 2023) package in R. As ordinal models will make the first fixed effect the model intercept (in the first model repeat "species"), the model iterations were repeated with "genus" as the intercept term so that all three pair-wise model estimates were obtained (species–genus, genus–family and species–family), thus allowing us to compare the effect size changes in sample completeness as taxonomic level increases. Furthermore, these paired effect sizes can be contrasted across the different spatial scales, thus allowing us to indirectly determine changes in sample completeness as spatial scale increases (QDGC < DGC < bioregion < biome).

It must be noted that with our procedure to quantify sample completeness there are multiple permutations that can result in the same outcome. An under-sampled

region may be the result of a region with few samples, even if the projected asymptote is reached (Fig. 1A), or as the result of a region with many samples, but whose asymptote is far from being reached (Fig. 1B). We do not try to differentiate between the sample or percentage completeness thresholds that are crossed, nor do we attribute different weights to these thresholds. These disparities between sample size and percentage completeness are larger at the lower sampling completeness levels, whereas at the higher sampling completeness levels, these disparities decrease, as well-sampled regions need to have many samples with an observed species richness within 10% of the estimated species richness (Fig. 1G).

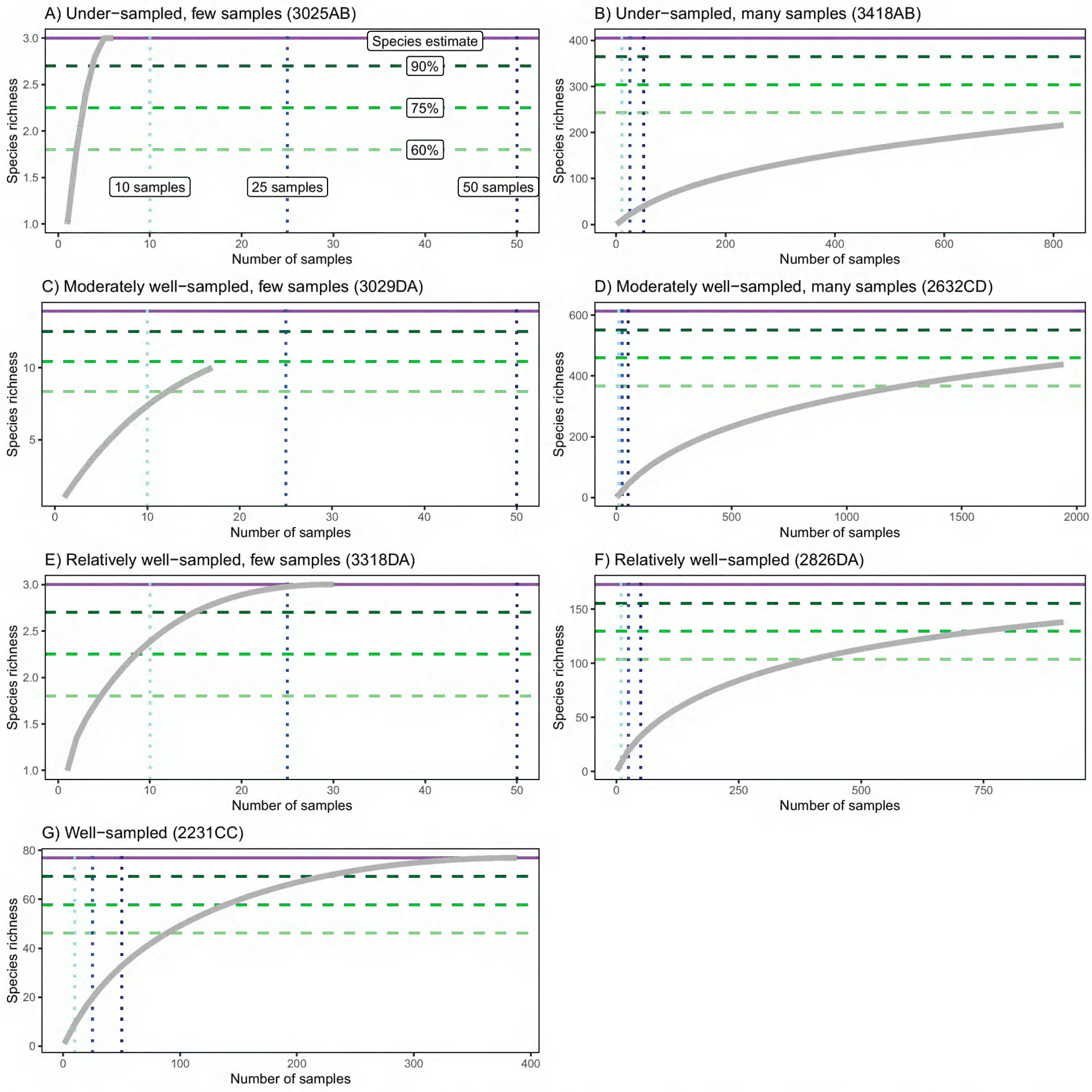

Figure 1. Example rarefaction curves (grey lines) for species level assemblages of select quarter degree grid cells. The horizontal purple line indicates the estimated species richness of the sample, while the dashed horizontal lines indicate the percentage completeness cut-off values, and the vertical dashed lines indicate the cut-off values for the accession numbers. For clarity, cut-off values are only annotated in plot **A**. Individual plots **A–G** show how different combinations of threshold values combine and result in under-sampled, moderately well-sampled, relatively well-sampled and well-sampled spatial units.

Results

Overall, 121 605 usable records were entered into our database (Table 1). Records identified to species level are the most predominant in the database, followed by genus- then family-level records (Table 1). When considering the completeness of our database, the species and genus records are relatively well- sampled, with 89.2% and 86.6% completeness percentage respectively, while the family records are well-sampled, with 100% completeness percentage calculated (Table 2).

The distribution of records across the country is somewhat uneven 41.3% of all QDGC in South Africa are lacking records, but this percentage decreases as the spatial scale increases, with only 5.3% of the DGCs lacking data, while neither the bioregions nor biomes are without spider records (Table 3). Spatially, records are distributed across the country, although there is an apparent concentration of records towards the eastern regions of South Africa, as well as along the coast from the eastern border with Mozambique, to north of Langebaan (Fig. 2). In the drier western regions of the country records are sparse, but where they do occur, they seemingly follow river courses. For example, the Orange River course is easily identifiable (Fig. 2). Consequently, spider diversity across species, genus and family level follows similar trends when considered

Table 1. Diversity summary of family, genus and species richness contained in each database, with total diversity counts shown in the last row. The last column indicates the number of records obtained from the individual databases.

Source database	Family richness	Genus richness	Species richness	Number of records
National Collection of Arachnida	72	557	1 718	73 649
Albany Museum	43	118	180	1 777
National Museum	65	404	643	16 061
KwaZulu-Natal Museum	66	342	680	10 517
Royal Museum of Central Africa	68	300	471	4 001
Iziko South African Museum	65	299	670	9 763
Ditsong National Museum of Natural History	66	256	266	5 837
Overall database	*74*	*639*	*2087*	*121 605*

Table 2. Sampling completeness of the South African spider fauna as a whole, the observed richness as well as estimated richness and standard error for each taxonomic level, as well as completeness percentages are shown.

Taxonomic level	Number of records	Observed richness	Estimated richness	SE	Completeness percentage (%)	Level of sampling completeness
Species	63 007	2086	2338.5	37.94	89.20	Relatively well
Genus	98 905	638	739	31.5	86.6	Relatively well
Family	121 605	74	74	0.6	100	Well

Table 3. Number of units per spatial scale, with the total number of sampled and unsampled units shown. Values in brackets indicate percentage of the total number either sampled or not.

Spatial scale	Total sampled	Not sampled	Total
Quarter degree grid cell	1172 (*58.7*)	823 (*41.3*)	1995
Degree grid cell	143 (*94.7*)	8 (*5.3*)	151
Bioregion	44 (*100*)	0	44
Biome	11 (*100*)	0	11

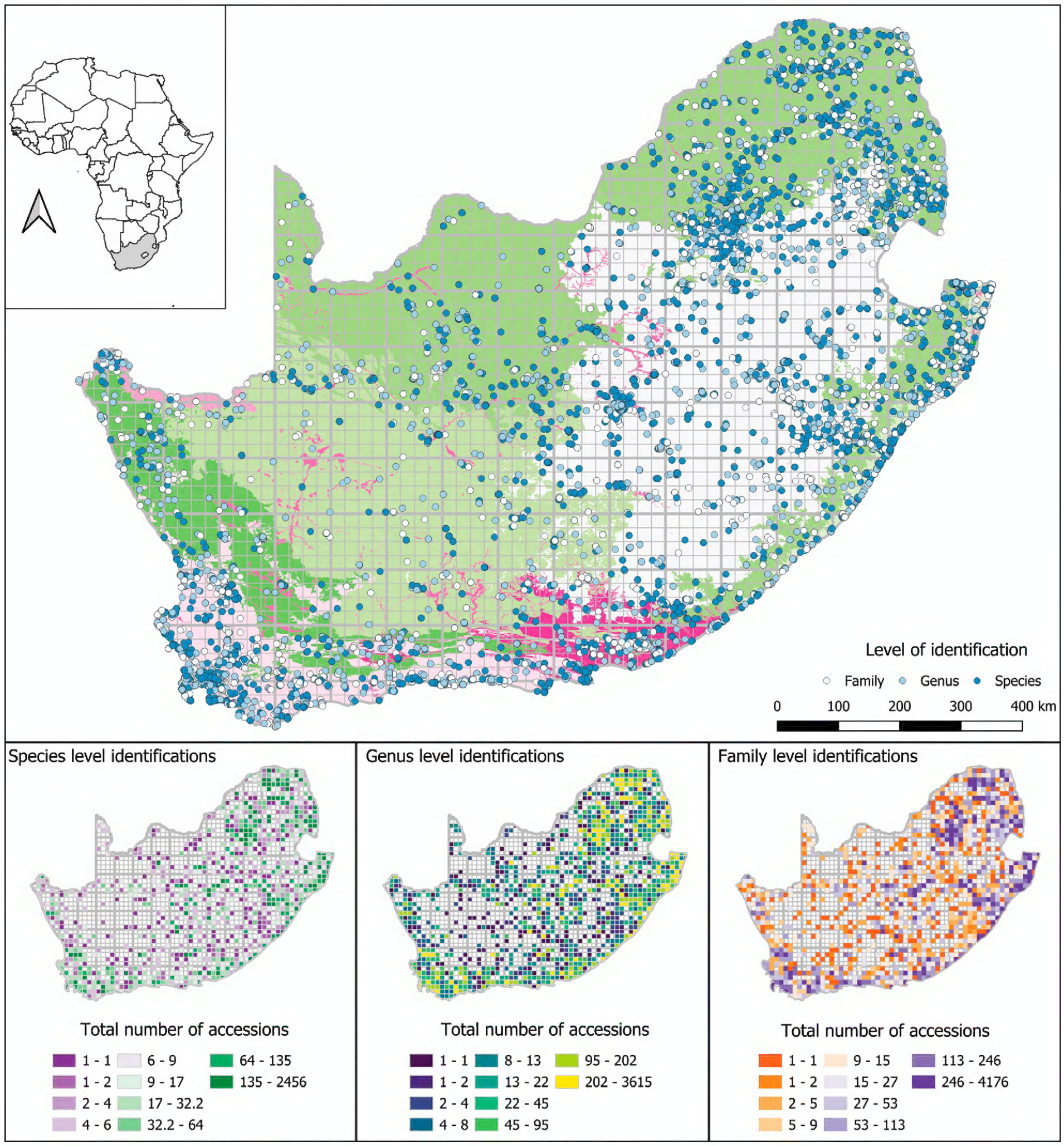

Figure 2. Distribution of the individual spider database records across South Africa. Colours indicate the lowest level individuals are identified to (family, genus or species level). For reference, the biomes are shown in green-pink fill. The map inserts show the total number of record accession in each Quarter Degree Grid Cell for records identified to family, genus and family levels.

across the four spatial levels (Fig. 3). Richness is highest in the eastern and coastal regions and lowest in the drier interior western regions of South Africa.

As we hypothesised, sample completeness scales up as both the spatial and taxonomic levels increase. Firstly, the most obvious of these completeness increases is when taxonomic scale is considered alone, at the same spatial scale (rows in Fig. 4). Here, the proportion of spatial units at higher completeness levels steadily increase as taxonomic level increases (Table 4, Fig. 4 rows). Secondly, as

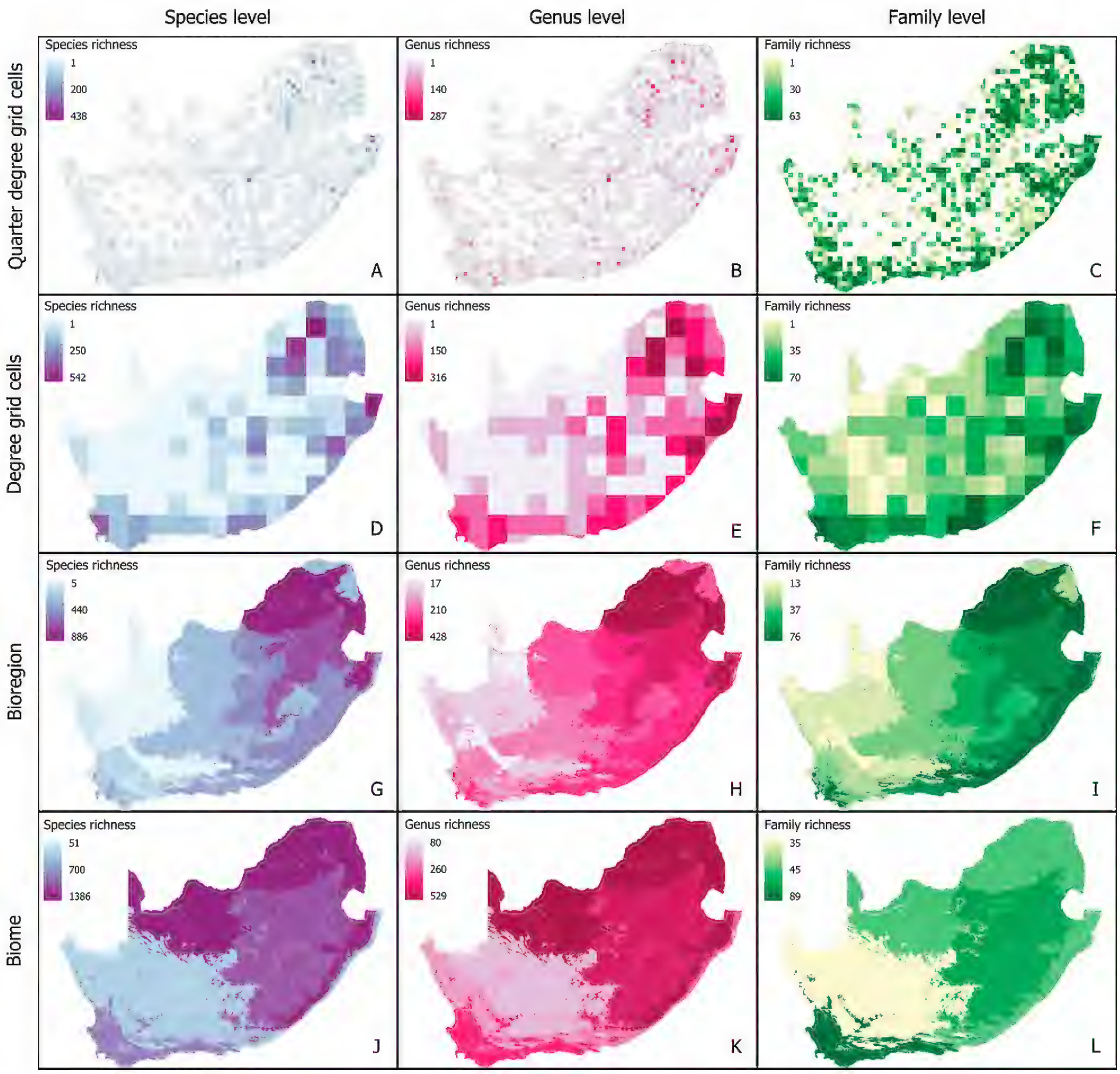

Figure 3. Distribution of spider diversity across the three taxonomic and four spatial levels considered. Darker colours indicate higher levels of diversity, with scales pertinent to each plot shown. Individual plots **A–L** show richness at all possible combinations of taxonomic and spatial scale.

spatial scale increases, but taxonomic scale remains constant (columns in Fig. 4), overall completeness increases. For example, spatial units that are well-sampled in terms of family level completeness increase from 3.7% to 33.77% to 47.72% to 63.64% when moving from the QDGC to biome level (Table 4, Fig. 4C, E, I, L). Although genus and species levels do not reach the well-sampled level at either the bioregion or biome scale, this pattern of increase holds true throughout the other levels of completeness (Table 4, Fig. 4). Finally, the combination of the increasing taxonomic and spatial scale results in the highest degree of sample completeness. At the lowest taxonomic and spatial scales, QDGC–species level, samples are the least complete, with only 0.15% of all QDGCs considered well-sampled and found in the northeastern and southern regions of the country (Table 4, Fig. 4A). As we hypothesized, sample completeness scales up as

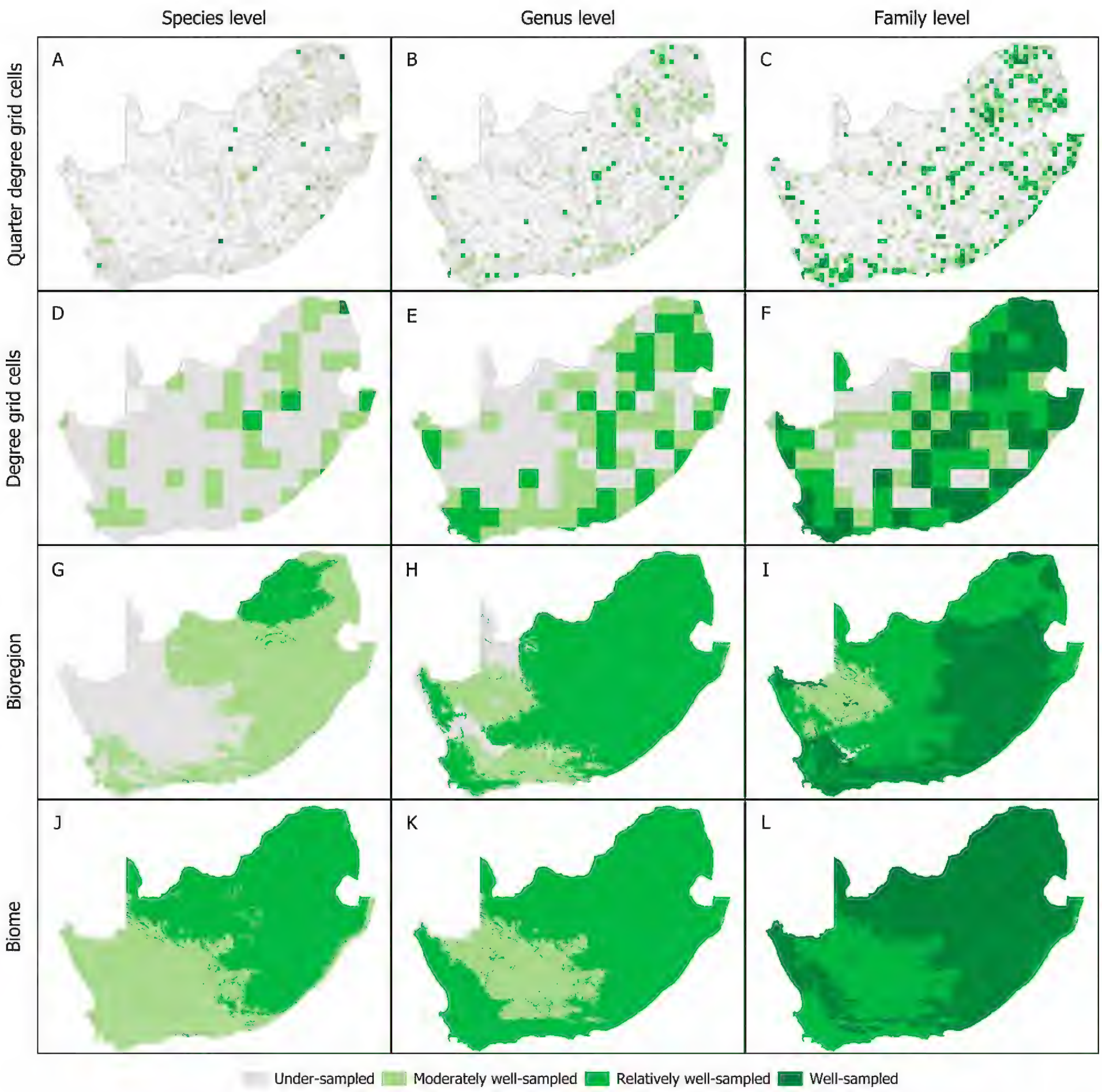

Figure 4. Changes in spider sampling completeness across both taxonomic level and spatial scale. Darker greens indicate a higher level of completeness per spatial unit. Individual plots **A–L** show sampling completeness at each combination of taxonomic and spatial scale.

both the spatial and taxonomic scales increase, such that at the highest scale combination, biome–family, sample completeness is the highest observed at 63.64% of all biomes being considered well-sampled (Table 4, Fig. 4L).

The CLMM results show that these positive changes in spider sample completeness are significant across all taxonomic and spatial levels (Table 5). The same patterns emerge both with and without the empty cells at the QDGC and DGC levels (Table 5, Fig. 4). When the magnitude and direction of the model estimates are plotted, it becomes evident that the level of spider sample completeness increases with spatial and taxonomic scale. When comparing the estimate size between genus to family (Fig. 5B) and species to family (Fig. 5C), there is a larger positive change in completeness than observed from species

Table 4. Distribution of well, relatively well, moderately well and under-sampled units across both the spatial and taxonomic scales considered. Numbers in brackets indicate the total percentage of units per criteria in relation to the total number of units in each spatial scale.

Spatial scale	Taxonomic scale	Well- sampled	Relatively well-sampled	Moderately well-sampled	Under-sampled	No data
Quarter degree grid cells	Species	3	8	81	890	1 013
	Genus	3	44	167	894	887
	Family	73	176	206	717	823
Degree grid cells	Species	1	4	36	99	11
	Genus	0	37	46	60	8
	Family	51	39	24	29	8
Bioregion	Species	0	2	17	25	
	Genus	0	23	9	12	-
	Family	21	16	5	2	
Biome	Species	0	3	6	2	
	Genus	0	8	2	1	-
	Family	7	3	1	0	

Table 5. Cumulative linked mixed effect model results Italics indicate significant comparisons. Rows with grey fill indicate models where empty cells were excluded.

Model iteration	n	Term comparison	Estimate	SE	z- value
Quarter degree grid cells (empty cells excluded)	3309	Species – Genus	1.73	0.19	*9.35*
		Genus – Family	2.64	0.19	*14.17*
		Species – Family	4.37	0.26	*16.62*
Quarter degree grid cells (empty cells included)	5985	Species – Genus	0.56	0.001	*527.4*
		Genus – Family	1.65	0.0001	*14 711*
		Species – Family	1.634	0.001	*1632.7*
Degree grid cells (empty cells excluded)	426	Species – Genus	2.57	0.34	*7.5*
		Genus – Family	3.63	0.37	*9.69*
		Species – Family	6.17	0.51	*12.04*
Degree grid cells (empty cells included)	453	Species – Genus	2.55	0.36	*7.12*
		Genus – Family	3.7	0.39	*9.48*
		Species – Family	6.25	0.57	*10.9*
Bioregion	132	Species – Genus	3.71	0.74	*5*
		Genus – Family	4.79	0.86	*5.56*
		Species – Family	8.51	1.33	*6.4*
Biome	33	Species – Genus	2.56	1.15	*2.22*
		Genus – Family	4.6	1.59	*2.9*
		Species – Family	7.18	2.09	*3.44*

to genus (Fig. 5A) at the QDGC and DGC levels. This, again, is what we hypothesised: the larger the difference in taxonomic scale, the larger the change in completeness. Furthermore, when comparing within taxonomic level combinations (Fig. 5A–C) individually, there is a positively larger change in completeness as spatial scale increases from QDGC to bioregion. At the biome level, this increase is not as apparent, but remains positive, which suggests that the chance of being considered at a higher level of sample completeness is still greater at these higher spatial scales.

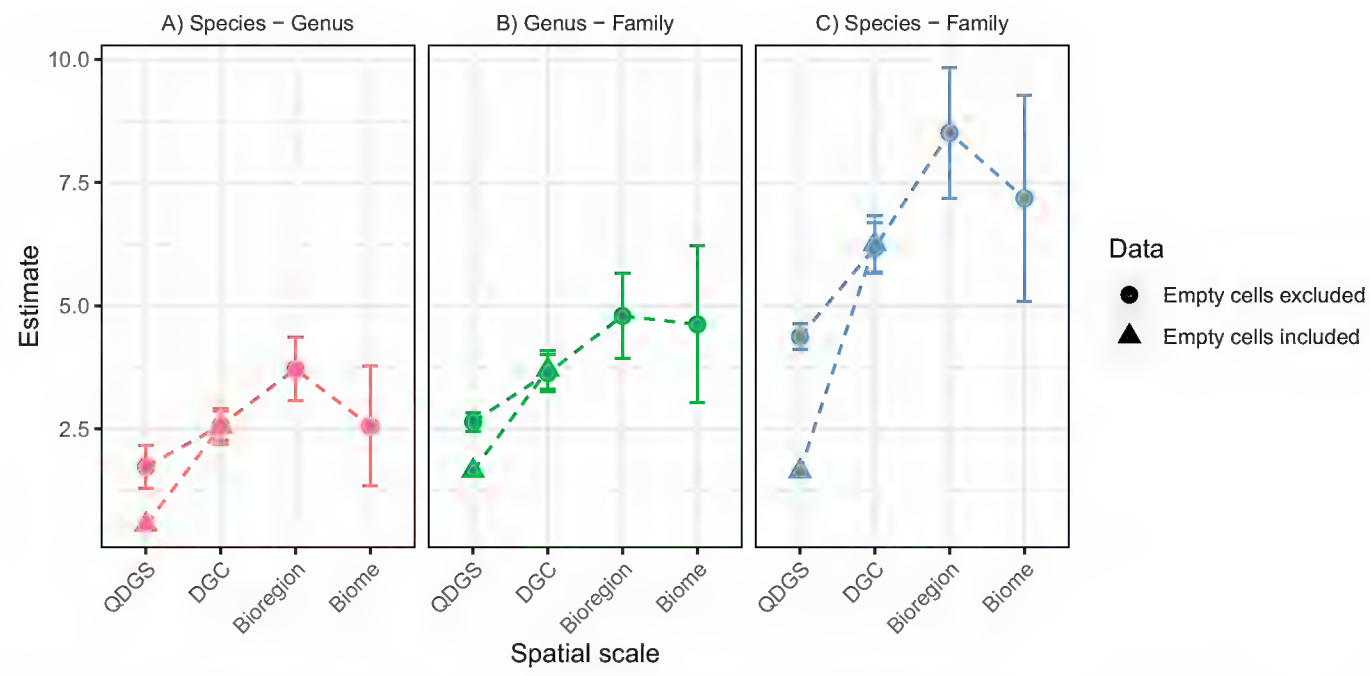

Figure 5. Cumulative linked mixed effect model estimates and standard errors shown for changes in spider sampling completeness as taxonomic level changes from **A** species to genus **B** genus to family and **C** species to family level. Within each panel, the direction and magnitude of the estimate change in spider sample completeness across spatial scale is shown. Model estimates for the quarter and degree grid cell levels with and without empty cells are also shown and indicated by the point shape.

Discussion

We show that, even for such an extensively studied taxon in South Africa, our fine-scale knowledge regarding spider assemblages in the country is relatively poor, and that extensive geographical bias exists within the database studied here. Although these biases exist, we have been able to identify regions within South Africa where spider assemblages can be considered complete, even at the finest scales considered. As scale becomes coarser, the overall completeness of the spider assemblages increases. Our demonstration that completeness of spider assemblages increases as both taxonomic and spatial scale increases is important, as it shows that the considerable amount of sampling that has been conducted on spiders in South Africa has been vital in determining a highly complete list of species. Interestingly, the estimated number of species within South Africa is 2338.5, which equates to 73 needing to be added for it to be considered complete based on the current rates of sampling and distribution of spider samples. However, based on the results of taxonomic revisions of Afrotropical spiders in recent decades (see World Spider Catalog 2024), particularly including South African taxa, many new species still await description, so this is itself a gross under-estimation of the country's spider diversity. This is exacerbated by the large parts of the country that remain unsampled.

Spider database

Spiders are mega-diverse in South Africa (Dippenaar-Schoeman et al. 2023), from which approximately 4% of the world's spiders have been recorded, with almost 59% of species considered country endemics (Dippenaar-Schoeman et al. 2023), although it is suggested that the lack of intensive sampling of spiders in neighbouring countries may artificially inflate the number of species considered as endemic (Foord et al. 2011a), which is a good example of a Wallacean shortfall within the database (Yang et al. 2013; Hortal et al. 2015). Although samples are incomplete in many areas, we have established a relatively com-

prehensive database of spatially referenced spider records from South Africa. There are fewer species represented here (2086) versus the 2265 species of the Checklist of Spiders (Dippenaar-Schoeman et al. 2023) given that these two studies draw from different data sources, differences in the total number of species can be expected. For example, the type material of the 187 spider species sampled by E. Simon between 1893 and 1910 are housed in overseas museums and were not included here. In addition to species-level records, our database contains records identified to genus or family level only, and as such, is representative of more spider genera than the Checklist (641 versus 495), this increase in numbers is due the new records being included after the publication of the checklist. The family diversity reported here represents 53% of the global family diversity of spiders (Foord et al. 2011a; World Spider Catalog 2024).

With regards to taxonomic accuracy, this database is likely the most accurate and relatively comprehensive, having combined the records of various curated natural history collections across the country. Nomenclature of the species level identification were verified so that species were not duplicated with old and new names following taxonomic transfers and synonymies. We do not include morphospecies identifications or pseudo-taxonomic records, and as such, diversity estimates are more likely to be a true reflection of the diversity at sites. Diversity estimates that are derived from morphospecies identifications are more likely to over- or under-estimate diversity, particularly in invertebrates, as most species-level identifications are based on obscure morphological features that are often overlooked when less trained individuals assign individual samples to morphospecies (Foord et al. 2013). In particular, the taxonomic resolution of the mega-diversity of spiders within South Africa is not available for other invertebrate taxa (Foord et al. 2011a).

Completeness

South Africa is a megadiverse country (Mittermeier et al. 2011; Colville et al. 2020), thus the finding of low levels of completeness in the spider assemblages is not surprising, nor is it a new hypothesis. Foord et al. (2011a) suggested that “South African spider systematics and ecology are in an exploratory phase” and further highlight biases in the distribution of records across the country. Our study further supports these findings, and in turn quantifies the completeness of the sampled spider communities. For example, 13 years ago Foord et al. (2011a) describe the Nama and Succulent Karoo as well as the Thicket biomes as poorly sampled in terms of species richness, while we in turn show that the two karoo biomes are moderately well-sampled and that the thicket is relatively well-sampled. This highlights the importance of maintaining up-to-date databases of species records, as well as the value of an integrated approach blending taxonomic and quantitative methodologies.

Abundance-based asymptotic estimators of species richness, such as Chao1 used here, are reliable estimators for species richness (Chao et al. 2014), with little variation in estimates when the sample grain size is reduced while maintaining area constant (Hortal et al. 2006). Even though we hold sample grain size constant across all scales by treating each spider record as a sample, the Chao1 estimate remains a feasible approach (Hortal et al. 2006; Chao and Jost 2012). At low sample sizes, Chao1 loses accuracy (Hortal et al. 2006), but with the inclusion

of the record thresholds here we avoid mis-classifying under-sampled regions as any higher class. Furthermore, it is well known that species richness correlates to sample size (Melo et al. 2003), yet this does not provide a species richness estimate, nor can it be used to calculate completeness of the spider assemblages.

Scale

Completeness of the spider assemblages is driven by scale. As taxonomic and geographic scale increase from fine to coarse scales, the overall completeness of the spider assemblages also increases. Considering that the increase in scale increases the size of the samples of each spatial unit, this is expected. What this highlights though, is that the larger scales absorb the sampling bias of the finer-scale samples. The western interior is a good example of this. At the finer scales, many QDGCs have no spider samples, but when looking at the completeness of the bioregions and biomes of this region, the spider assemblage is more often than not considered as moderately well-sampled. This does not mean that these determined completeness levels of the higher spatial and taxonomic scales are incorrect, but rather that completeness cannot be interpreted in the reverse direction. For example, the moderately well-sampled Nama Karoo is linked to the biome scale, but this completeness category cannot be applied to all the QDGC grid cells that fall within the region of the Nama Karoo, as many of these grid cells do not have any reported spider samples.

The difference between the estimated and observed species richness is the number of not yet sampled species (Chao et al. 2014). With regards to knowledge gaps, these “missing” species are not necessarily undescribed species (i.e. Linnean shortfall), but also species that are under-sampled and whose ranges are poorly understood (i.e. a Wallacean shortfall) (Bini et al. 2006; Hortal et al. 2015; Assis 2018; Diniz-Filho et al. 2023). Thus, as completeness of the spider knowledge base increases with spatial and taxonomic scales, the size of both the Linnean and Wallacean gaps incrementally decreases, such that at large spatial (biome) and taxonomic (family) scales, the number of families not yet sampled are all less than 40% of the estimated family diversity, with more biomes being represented by a unknown proportion of less than 25% (inverses of the threshold cutoff values).

It must be noted, though, that the estimates of richness depend on the input data and will constantly change as more and more samples are added (Chao and Jost 2012; Chao et al. 2020; Kusumoto et al. 2023). Standard errors are calculated for each richness estimate, and here we have used the mean estimate to determine the sample completeness. Thus, there is inherent variation in these estimates, and they are not a singular and perfect estimate of richness. This point is exemplified by the fact that our total species richness observed here is less than that of the Spider Checklist, yet our estimated species richness is greater than that of the Spider Checklist. Given that this study draws on a different set of data than that of the Spider Checklist, differences in observed as well as expected richness will vary, as each data source would contain a different cohort of species. However, the variation in estimates and richness should not detract from the fact that here we have shown explicitly how the knowledge gaps within the South African spiders vary across both spatial and taxonomic scales. Even though at larger scales, our understanding of spider

distributions and occurrences across the country is relatively incomplete given the shear diversity of spiders in the country, the investment in sampling over the 300 years of spider research in the country has been hugely successful and cannot be overlooked. However, much still needs to be done to remedy the under-sampling in the western half of the country, particularly, to improve species-level distribution data, and taxonomic inputs to relieve the considerable Linnean shortfall. A suggestion, going forward and considering both limited funding and time, would be to systematically sample at the centroid of each DGC with little to no samples in the interior western regions to fill in these gaps in species-level distributions. Sampling at the QDGC would be unadvisable as there would be too many sites to feasibly sample in the short term, while deciding on sampling sites at the bioregion and biome region will not capture finer scale variation in species-level distributions. Collecting using rapid sampling protocols has been shown to generate large numbers of specimens and species-level records in sampling intervals less than a week (Haddad and Dippenaar-Schoeman 2015; Booysen and Haddad 2021; Haddad 2021), so applying this approach would enable to generation of sizable datasets in the under-sampled DGCs using the limited human resources available.

Bias

The distribution of spider records within South Africa is congruent with that of many invertebrate groups within South Africa, such as dragonflies (records from all families in the order) (Simaika and Samways 2009; Basel et al. 2021; 2024), dung beetles (records from all families in the order) (Davis and Scholtz 2020), and katydids (records only from the Tettigoniidae family)(Bazelet et al. 2016), where records are concentrated along the coastal regions of the country, and comparatively fewer in the interior drier regions of South Africa. For all these invertebrate groups, ecological drivers are suggested as the reason for the diversity distribution, although no mention of sampling bias is made. Spider records were under-sampled in the Northern Cape and North-West province of South Africa (Foord et al. 2020) and remain so here. As evidenced here by the large number of QDGC grids still unsampled, which remain concentrated in the Northern Cape and North-West provinces even though at higher scales these regions appear to be moderately, relatively and well-sampled (Fig. 3).

Bias in sample completeness will also scale with taxonomic and spatial scale of the study in question. Bazelet et al. (2016) compared katydid diversity between biodiversity hotspots and non-hotspots in South Africa and considered katydid assemblages as complete, having constructed and compared two accumulation curves for hotspots and non-hotspots. They did not consider finer scales across South Africa. Notably, there are an estimated 169 species of katydid known in South Africa (Thompson et al. 2017), so a large-scale analysis (hotspot versus non-hotspot) of katydids is likely to quickly reach an asymptote. Conversely, here we have a list of more than 2000 species, and at the smaller spatial scales sample completeness is very low. The lower spider species completeness levels at the bioregion and biome levels are justifiable when compared to the coarse level of Bazelet et al. (2016), as the spider diversity here is of an order magnitude greater than the diversity of katydids considered, as the katydids represent a single family only.

We have not distinguished between sample origins or sampling methods, but rather treated each record equally. Systematic and opportunistic surveys and sampling methods will lead to different numbers of samples. Given that the records here span a range of collection trips, the duration, methods and approaches employed will differ markedly. Record keeping between institutions and individuals will also result in bias within datasets, particularly in cases where taxonomic expertise is available to improve the resolution of identifications and keep this updated in line with global taxonomic changes (World Spider Catalog 2024). Here we set out stringent requirements for data to be retained in order to minimize record keeping errors in our dataset.

Conclusions and implications

Spider assemblage completeness is a direct result of both the spatial and taxonomic scales being considered. Furthermore, the scaling of completeness can only be interpreted in one direction, from fine to coarse and not the other way around. As scale increases, so too does the overall completeness of the spider assemblages. This will have important implications for future spider research and conservation. Given that the regions where completeness is highest across all scales correspond strongly to metropolitan areas and the areas with the highest threats to biodiversity in South Africa, and that there is a notable global decline in insect and invertebrate diversity (Cardoso et al. 2020), the determination of trends in invertebrate diversity across regions and at different scales is of paramount importance. Without understanding the underlying patterns of diversity and distributions, conservation efforts are likely to be ineffective.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical clearance required as this study is based on existing museum collections.

Funding

AvdM was funded through her postdoctoral host's (Daryl Codron) National Research Foundation Competitive Grant for Rated Researched (UID 137968). CRH and SHF were funded through the National Research Foundations Incentive Funding for Rated Researchers (grants #132687 and #87311 respectively).

Author contributions

AvdM conceptualized the study, performed data collection and cleaning, as well as the analyses and wrote the first draft of the manuscript. CRH collected and cleaned data, as well as verified species identification and contributed to the first draft of the manuscript. ASD performed identifications and established the NCA database from which much of the data originated, as well as contributing to the first draft of the manuscript. SHF was the inspiration for the study having provided over half the database freely and without hesitation, having compiled the SANSA database, his input in the final draft of the manuscript was sorely missed.

Author ORCIDs

Aileen C. van der Mescht https://orcid.org/0000-0003-3849-8636
Charles R. Haddad https://orcid.org/0000-0002-2317-7760
Stefan H. Foord https://orcid.org/0000-0002-9195-2562
Ansie S. Dippenaar-Schoeman https://orcid.org/0000-0003-1532-1379

Data availability

The full dataset is not publicly available, but will be made available upon reasonable request.

References

Aceves-Bueno E, Adeleye AS, Feraud M, Huang Y, Tao M, Yang Y, Anderson SE (2017) The accuracy of citizen science data: A quantitative review. Source. Bulletin of the Ecological Society of America 98(4): 278–290. https://doi.org/10.1002/bes2.1336

Alves-Martins F, Stropp J, Juen L, Ladle RJ, Lobo JM, Martinez-Arribas J, Júnior PDM, Brasil LS, Ferreira VRS, Bastos RC, Córdoba-Aguilar A, Medina-Espinoza EF, Dutra S, Vilela DS, Cordero-Rivera A, del Palacio A, Ramírez A, Carvalho-Soares AA, Farias ABS, de Resende BO, dos Santos B, Bota-Sierra CA, Mendoza-Penagos CC, Veras DS, Anjos-Santos D, Périco E, González-Soriano E, de Oliveira Roque F, Lozano F, de Carvalho FG, Lencioni FAA, Palacino-Rodríguez F, Ortega-Salas H, Venâncio H, Sanmartín-Villar I, Muzón J, Santos JC, Montes-Fontalvo J, da Silva Brito J, da Silva Pereira JL, Oliveira-Junior JMB, Dias-Silva K, Ferreira KG, Calvão LB, Pérez-Gutiérrez LA, Rodrigues ME, Dalzochio MS, Rocha-Ortega M, von Ellenrieder N, Hamada N, Pessacq P, Rodríguez P, Martins RT, Guillermo-Ferreira R, Koroiva R, Miguel TB, Mendes TP, Neiss UG, de Almeida WR, Hortal J (2024) Sampling completeness changes perceptions of continental scale climate–species richness relationships in odonates. Journal of Biogeography 00(7): 1–15. https://doi.org/10.1111/jbi.14810

Assis LCS (2018) Revisiting the Darwinian shortfall in biodiversity conservation. Biodiversity and Conservation 27(11): 2859–2875. https://doi.org/10.1007/s10531-018-1573-3

Banks-Leite C, Betts MG, Ewers RM, Orme CDL, Pigot AL (2022) The macroecology of landscape ecology. Trends in Ecology & Evolution 37(6): 480–487. https://doi.org/10.1016/j.tree.2022.01.005

Basel AM, Simaika JP, Samways MJ, Midgley GF, MacFadyen S, Hui C (2021) Assemblage reorganization of South African dragonflies due to climate change. Diversity & Distributions 27(12): 2542–2558. https://doi.org/10.1111/ddi.13422

Basel AM, Simaika JP, Samways MJ, Midgley GF, Latombe G, MacFadyen S, Hui C (2024) Drivers of compositional turnover in narrow-ranged and widespread dragonflies and damselflies in Africa. Insect Conservation and Diversity 17(3): 1–11. https://doi.org/10.1111/icad.12718

Bazelet CS, Thompson AC, Naskrecki P (2016) Testing the efficacy of global biodiversity hotspots for insect conservation: The case of South African katydids. PLoS ONE 11(9): 1–17. https://doi.org/10.1371/journal.pone.0160630

Bini LM, Diniz-Filho JAF, Rangel TFLVB, Bastos RP, Pinto MP (2006) Challenging Wallacean and Linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. Diversity & Distributions 12(5): 475–482. https://doi.org/10.1111/j.1366-9516.2006.00286.x

Booysen R, Haddad CH (2021) Season determines the efficiency of a rapid sampling protocol for non-acarine arachnids (Chelicerata: Arachnida) in Afrotemperate grassland biotopes. Austral Entomology 60(4): 682–697. https://doi.org/10.1111/aen.12573

Burley HM, Mokany K, Ferrier S, Laffan SW, Williams KJ, Harwood TD (2016) Macroecological scale effects of biodiversity on ecosystem functions under environmental change. Ecology and Evolution 6(8): 2579–2593. https://doi.org/10.1002/ece3.2036

Callaghan CT, Ozeroff I, Hitchcock C, Chandler M (2020) Capitalizing on opportunistic citizen science data to monitor urban biodiversity: A multi-taxa framework. Biological Conservation 251: 108753. https://doi.org/10.1016/j.biocon.2020.108753

Cardoso P, Erwin TL, Borges PAV, New TR (2011) The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation 144(11): 2647–2655. https://doi.org/10.1016/j.biocon.2011.07.024

Cardoso P, Barton PS, Birkhofer K, Chichorro F, Deacon C, Fartmann T, Fukushima CS, Gaigher R, Habel JC, Hallmann CA, Hill MJ, Hochkirch A, Kwak ML, Mammola S, Ari Noriega J, Orfinger AB, Pedraza F, Pryke JS, Roque FO, Settele J, Simaika JP, Stork NE, Suhling F, Vorster C, Samways MJ (2020) Scientists' warning to humanity on insect extinctions. Biological Conservation 242: 108426. https://doi.org/10.1016/j.biocon.2020.108426

Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. Ecology 93(12): 2533–2547. https://doi.org/10.1890/11-1952.1

Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell RK, Ellison AM (2014) Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. Ecological Monographs 84(1): 45–67. https://doi.org/10.1890/13-0133.1

Chao A, Kubota Y, Zelený D, Chiu CH, Li CF, Kusumoto B, Yasuhara M, Thorn S, Wei CL, Costello MJ, Colwell RK (2020) Quantifying sample completeness and comparing diversities among assemblages. Ecological Research 35(2): 292–314. https://doi.org/10.1111/1440-1703.12102

Christensen R (2023) ordinal - Regression Models for Ordinal Data.

Clough Y, Kruess A, Kleijn D, Tscharntke T (2005) Spider diversity in cereal fields: Comparing factors at local, landscape and regional scales. Journal of Biogeography 32(11): 2007–2014. https://doi.org/10.1111/j.1365-2699.2005.01367.x

Colville JF, Beale CM, Forest F, Altwegg R, Huntley B, Cowling RM, Ackerly DD, Latimer AM, Proches S (2020) Plant richness, turnover, and evolutionary diversity track gradients of stability and ecological opportunity in a megadiversity center. Proceedings of the National Academy of Sciences of the United States of America 117(33): 20027–20037. https://doi.org/10.1073/pnas.1915646117

Cornford R, Deinet S, De Palma A, Hill SLL, McRae L, Pettit B, Marconi V, Purvis A, Freeman R (2021) Fast, scalable, and automated identification of articles for biodiversity and macroecological datasets. Global Ecology and Biogeography 30(1): 339–347. https://doi.org/10.1111/geb.13219

Davis ALV, Scholtz CH (2020) Dung beetle conservation biogeography in southern Africa: Current challenges and potential effects of climatic change. Biodiversity and Conservation 29(3): 667–693. https://doi.org/10.1007/s10531-019-01904-7

De Mas E, Chust G, Pretus JL, Ribera C (2009) Spatial modelling of spider biodiversity: Matters of scale. Biodiversity and Conservation 18(7): 1945–1962. https://doi.org/10.1007/s10531-008-9566-2

Diniz-Filho JAF, Jardim L, Guedes JJM, Meyer L, Stropp J, Frateles LEF, Pinto RB, Lohmann LG, Tessarolo G, De Carvalho CJB, Ladle RJ, Hortal J (2023) Macroecological links between the Linnean,Wallacean, and Darwinian shortfalls. Frontiers of Biogeography 15(2): e59566. https://doi.org/10.21425/F5FBG59566

Dippenaar-Schoeman AS, Haddad CR, Foord S, Lyle R, Lotz L, Helberg L, Mathebula S, Van Den Berg A, Marais P, Van Den Berg AM, Van Niekerk E, Jocqué R (2010) The First Atlas of the Spiders of South Africa (Arachnida: Araneae). South African National Survey of Arachnida Technical Report. https://doi.org/10.5281/zenodo.7628809 [October 8, 2024]

Dippenaar-Schoeman AS, Haddad CR, Foord S, Lyle R, Lotz L, Marais P (2015) South African National Survey of Arachnida (SANSA): Review of Current Knowledge, Constraints and Future Needs For Documenting Spider Diversity (Arachnida: Araneae). Transaction of the Royal Society of South Africa 70(3): 245–275. htpps://doi.org/10.1080/0035919X.2015.1088486

Dippenaar-Schoeman A, Lyle R, van den Berg A (2012) Bioinformatics on the spiders of South Africa. Serket = Sarkat 13: 121–127.

Dippenaar-Schoeman AS, Haddad CR, Lotz LN, Booysen R, Steenkamp RC, Foord SH (2023) Checklist of the spiders (Araneae) of South Africa. African Invertebrates 64(3): 221–289. https://doi.org/10.3897/AfrInvertebr.64.111047

Drinkrow DR, Cherry MI, Siegfried WR (1994) The role of natural history museums in preserving biodiversity in South Africa. South African Journal of Science 90: 470–479.

Foord SH, Dippenaar-Schoeman AS (2016) The effect of elevation and time on mountain spider diversity: A view of two aspects in the Cederberg mountains of South Africa. Journal of Biogeography 43(12): 2354–2365. https://doi.org/10.1111/jbi.12817

Foord SH, Mafadza MM, Dippenaar-Schoeman AS, Van Rensburg BJ (2008) Micro-scale heterogeneity of spiders (Arachnida:Araneae) in the Soutpansberg, South Africa: a comparative survey and inventory in representative habitats. African Zoology 43(2): 156–174. https://doi.org/10.1080/15627020.2008.11657233

Foord S, Dippenaar-Schoeman A, Haddad C (2011a) South African spider diversity: African perspectives on the conservation of a Mega-diverse group. In: Grillo O, Venora G (Eds) Changing diversity in changing environment. INTECH Open Access Publisher, Rijeka,Croatia, 163–182.

Foord SF, Dippenaar-Schoeman AS, Haddad CR, Lotz LN, Lyle R (2011b) The faunistic diversity of spiders (Arachnida: Araneae) of the savanna biome in South Africa. Transactions of the Royal Society of South Africa 66(3): 170–201. https://doi.org/10.1080/0035919X.2011.639406

Foord SH, Dippenaar-Schoeman AS, Stam EM (2013) Surrogates of spider diversity, leveraging the conservation of a poorly known group in the Savanna Biome of South Africa. Biological Conservation 161: 203–212. https://doi.org/10.1016/j.biocon.2013.02.011

Foord SH, Dippenaar-Schoeman AS, Haddad CR, Lyle R, Lotz LN, Sethusa T, Raimondo D (2020) The South African National Red List of spiders: Patterns, threats, and conservation. The Journal of Arachnology 48(2): 110–118. https://doi.org/10.1636/0161-8202-48.2.110

Fraisl D, Hager G, Bedessem B, Gold M, Hsing PY, Danielsen F, Hitchcock CB, Hulbert JM, Piera J, Spiers H, Thiel M, Haklay M (2022) Citizen science in environmental and ecological sciences. Nature Reviews. Methods Primers 2(1): 64. https://doi.org/10.1038/s43586-022-00144-4

Garretson A, Cuddy T, Duffy AG, Forkner RE (2023) Citizen science data reveal regional heterogeneity in phenological response to climate in the large milkweed bug, Oncopeltus fasciatus. Ecology and Evolution 13(7): e10213. https://doi.org/10.1002/ece3.10213

Gómez-Rodríguez C, Baselga A (2018) Variation among European beetle taxa in patterns of distance decay of similarity suggests a major role of dispersal processes. Ecography 41(11): 1825–1834. https://doi.org/10.1111/ecog.03693

Gonçalves-Souza T, Diniz-Filho JAF, Romero GQ (2014) Disentangling the phylogenetic and ecological components of spider phenotypic variation. PLoS ONE 9(2): e89314. https://doi.org/10.1371/journal.pone.0089314

Haddad CR (2021) Undergraduate entomology field excursions are a valuable source of biodiversity data: A case study for spider (Araneae) bycatches in ecological studies. Biodiversity and Conservation 30(14): 4199–4222. https://doi.org/10.1007/s10531-021-02301-9

Haddad CR, Dippenaar-Schoeman AS (2015) Diversity of non-acarine arachnids of the Ophathe Game Reserve, South Africa: Testing a rapid sampling protocol. Koedoe 57(1): 1–15. https://doi.org/10.4102/koedoe.v57i1.1255

Haddad CR, Dippenaar-Schoeman AS, Lyle R, Foord SH, Lotz LN (2013) The faunistic diversity of spiders (Arachnida: Araneae) of the South African Grassland Biome. Transactions of the Royal Society of South Africa 68(2): 97–122. https://doi.org/10.1080/0035919X.2013.773267

Haddad CR, de Jager LJC, Foord SH (2019) Habitats and cardinal directions are key variables structuring spider leaf litter assemblages under Searsia lancea. Pedobiologia 73: 10–19. https://doi.org/10.1016/j.pedobi.2019.01.002

Hamer M (2012) An assessment of zoological research collections in South Africa. South African Journal of Science 108(11/12): 1–11. https://doi.org/10.4102/sajs.v108i11/12.1090

Hortal J, Borges PAV, Gaspar C (2006) Evaluating the performance of species richness estimators: Sensitivity to sample grain size. Journal of Animal Ecology 75(1): 274–287. https://doi.org/10.1111/j.1365-2656.2006.01048.x

Hortal J, De Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics 46(1): 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hsiech TC, Ma KH, Chao A (2020) iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.20. http://chao.stat.nthu.edu.tw/wordpress/software-download/

Jacobs C, Zipf A (2017) Completeness of citizen science biodiversity data from a volunteered geographic information perspective. Geo-Spatial Information Science 20(1): 3–13. https://doi.org/10.1080/10095020.2017.1288424

Janion-Scheepers C, Measey J, Braschler B, Chown SL, Coetzee L, Colville JF, Dames J, Davies AB, Davies SJ, Davis ALV, Dippenaar-Schoeman AS, Duffy GA, Fourie D, Griffiths C, Haddad CR, Hamer M, Herbert DG, Hugo-Coetzee EA, Jacobs A, Jacobs K, Rensburg CJ, van, Lamani S, Lotz LN (2016) Soil biota in a megadiverse country: Current knowledge and future research directions in South Africa. Pedobiologia 59(3): 129–174. https://doi.org/10.1016/j.pedobi.2016.03.004

Jiménez-Valverde A, Lobo JM (2007) Determinants of local spider (Araneidae and Thomisidae) species richness on a regional scale: Climate and altitude vs. habitat structure. Ecological Entomology 32(1): 113–122. https://doi.org/10.1111/j.1365-2311.2006.00848.x

Joseph GS, Mauda EV, Seymour CL, Munyai TC, Dippenaar-Schoeman A, Foord SH (2018) Landuse Change in Savannas Disproportionately Reduces Functional Diversity of Invertebrate Predators at the Highest Trophic Levels: Spiders as an Example. Ecosystems 21(5): 930–942. https://doi.org/10.1007/s10021-017-0194-0

Kusumoto B, Chao A, Eiserhardt WL, Svenning J-C, Shiono T, Kubota Y (2023) Occurrence-based diversity estimation reveals macroecological and conservation knowledge gaps for global woody plants. Science Advances, eadh9719. https://www.science.org

Martín-Devasa R, Jiménez-Valverde A, Leprieur F, Baselga A, Gómez-Rodríguez C (2024) Dispersal limitation shapes distance-decay patterns of European spiders at the continental scale. Global Ecology and Biogeography 33(4): e13810. https://doi.org/10.1111/geb.13810

Melo AS, S Pereira RA, Santos AJ, Shepherd GJ, Machado G, et al. (2003) Comparing species richness among assemblages using sample units: why not use extrapolation methods to standardize different sample sizes? IKOS 101: 398–410. https://doi.org/10.1034/j.1600-0706.2003.11893.x

Mittermeier RA, Turner WR, Larsen FW, Brooks TM, Gascon C (2011) Global Biodiversity Conservation: The Critical Role of Hotspots. In: Biodiversity Hotspots. Springer Berlin Heidelberg, Berlin, Heidelberg, 3–22. https://doi.org/10.1007/978-3-642-20992-5_1

Mora C, Tittensor DP, Myers RA (2008) The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. Proceedings. Biological Sciences 275(1631): 149–155. https://doi.org/10.1098/rspb.2007.1315

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403(6772): 853–858. https://doi.org/10.1038/35002501

Oliveira U, Paglia AP, Brescovit AD, de Carvalho CJB, Silva DP, Rezende DT, Leite FSF, Batista JAN, Barbosa JPPP, Stehmann JR, Ascher JS, de Vasconcelos MF, De Marco Jr P, Löwenberg-Neto P, Dias PG, Ferro VG, Santos AJ (2016) The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Diversity & Distributions 22(12): 1232–1244. https://doi.org/10.1111/ddi.12489

Pärtel M, Bennett JA, Zobel M (2016) Macroecology of biodiversity: Disentangling local and regional effects. The New Phytologist 211(2): 404–410. https://doi.org/10.1111/nph.13943

Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10(1): 439. https://doi.org/10.32614/RJ-2018-009

QGIS Development Team (2020) QGIS Geographic. Information Systems.

Qian H (2020) Are species lists derived from modeled species range maps appropriate for macroecological studies? A case study on data from BIEN. Basic and Applied Ecology 48: 146–156. https://doi.org/10.1016/j.baae.2020.08.003

R Core Team (2020) R: a language and environment for statistical computing.

Ramírez F, Sbragaglia V, Soacha K, Coll M, Piera J (2022) Challenges for Marine Ecological Assessments: Completeness of Findable, Accessible, Interoperable, and Reusable Biodiversity Data in European Seas. Frontiers in Marine Science 8: 802235. https://doi.org/10.3389/fmars.2021.802235

Robertson MP, Cumming GS, Erasmus BFN (2010) Getting the most out of atlas data. Diversity & Distributions 16(3): 363–375. https://doi.org/10.1111/j.1472-4642.2010.00639.x

Rutherford MC, Mucina L, Powrie LW (2005) Biomes and bioregions of Southern Africa. Vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, Pretoria.

Sánchez-Fernández D, Yela JL, Acosta R, Bonada N, García-Barros E, Guisande C, Heine J, Millán A, Munguira ML, Romo H, Zamora-Muñoz C, Lobo JM (2022) Are patterns of sampling effort and completeness of inventories congruent? A test using databases for five insect taxa in the Iberian Peninsula. Insect Conservation and Diversity 15(4): 406–415. https://doi.org/10.1111/icad.12566

Schmidt MH, Thies C, Nentwig W, Tscharntke T (2008) Contrasting responses of arable spiders to the landscape matrix at different spatial scales. Journal of Biogeography 35(1): 157–166. https://doi.org/10.1111/j.1365-2699.2007.01774.x

Scoble MJ (2010) Natural history collections digitization: Rationale and value. Biodiversity Informatics 7: 77–80. http://data.gbif.org. https://doi.org/10.17161/bi.v7i2.3994

Simaika JP, Samways MJ (2009) Reserve selection using Red Listed taxa in three global biodiversity hotspots: Dragonflies in South Africa. Biological Conservation 142(3): 638–651. https://doi.org/10.1016/j.biocon.2008.11.012

South African National Biodiversity Institute (2006) The vegetation map of South Africa, Lesotho and Swaziland.

Teng SN, Svenning J-C, Santana J, Reino L, Abades S, Xu C (2020) Linking Landscape Ecology and Macroecology by Scaling Biodiversity in Space and Time. Current Landscape Ecology Reports 5(2): 25–34. https://doi.org/10.1007/s40823-020-00050-z

Thompson AC, Bazelet CS, Naskrecki P, Samways MJ (2017) Adapting the dragonfly biotic index to a katydid (tettigoniidae) rapid assessment technique: Case study of a biodiversity hotspot, the Cape Floristic Region, South Africa. Journal of Orthoptera Research 26: 63–71. https://doi.org/10.3897/jor.26.14552

Troia MJ, McManamay RA (2016) Filling in the GAPS: Evaluating completeness and coverage of open-access biodiversity databases in the United States. Ecology and Evolution 6(14): 4654–4669. https://doi.org/10.1002/ece3.2225

Vergara-Asenjo G, Alfaro FM, Pizarro-Araya J (2023) Linnean and Wallacean shortfalls in the knowledge of arthropod species in Chile: Challenges and implications for regional conservation. Biological Conservation 281: 110027. https://doi.org/10.1016/j.biocon.2023.110027

Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation biogeography: Assessment and prospect. Diversity & Distributions 11(1): 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Wilkinson BH, Ivany LC, Drummond CN (2021) Estimating vertebrate biodiversity using the tempo of taxonomy-a view from Hubbert's peak. Biological Journal of the Linnean Society. Linnean Society of London 134(2): 402–422. www.mammaldiversity.org. https://doi.org/10.1093/biolinnean/blab080

Wilson JD, Bond JE, Harvey MS, Ramírez MJ, Rix MG (2023) Correlation with a limited set of behavioral niches explains the convergence of somatic morphology in mygalomorph spiders. Ecology and Evolution 13(1): e9706. https://doi.org/10.1002/ece3.9706

Wolf S, Mahecha MD, Sabatini FM, Wirth C, Bruelheide H, Kattge J, Moreno Martínez Á, Mora K, Kattenborn T (2022) Citizen science plant observations encode global trait patterns. Nature Ecology & Evolution 6(12): 1850–1859. https://doi.org/10.1038/s41559-022-01904-x

World Spider Catalog (2024) World Spider Catalog.

Wüest RO, Zimmermann NE, Zurell D, Alexander JM, Fritz SA, Hof C, Kreft H, Normand S, Cabral JS, Szekely E, Thuiller W, Wikelski M, Karger DN (2020) Macroecology in the age of Big Data – Where to go from here? Journal of Biogeography 47(1): 1–12. https://doi.org/10.1111/jbi.13633

Yang W, Ma K, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness-environment models. Journal of Biogeography 40(8): 1415–1426. https://doi.org/10.1111/jbi.12108